Atmospheric Environment 122 (2015) 748–762

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

An adaptive Bayesian inference algorithm to estimate the parameters of a hazardous atmospheric release

Harizo Rajaona [a, b, *], François Septier [a], Patrick Armand [b], Yves Delignon [a], Christophe Olry [c], Armand Albergel [c], Jacques Moussafir [c]

[a] Institut Mines-Télécom/Télécom Lille/LAGIS UMR CNRS 8219, 59650 Villeneuve-d'Ascq, France
[b] CEA-DAM, DIF, F-91297 Arpajon, France
[c] ARIA Technologies, 8 rue de la Ferme, 92100 Boulogne-Billancourt, France

HIGHLIGHTS

- We tackle the problem of estimating the location and the distribution of an atmospheric polluting source.
- We use Bayesian inference as a main framework, to be able to quantify the uncertainty.
- We derive an analytical way to estimate the release-rate by exploiting Gaussian assumptions on its prior distribution.
- We estimate the location by using an adaptive Monte Carlo algorithm based on the principles of Importance Sampling.
- We test our algorithm on a combination of synthetic and experimental data taken from the FFT07 experiment.

ARTICLE INFO

Article history:
Received 9 January 2015
Received in revised form
8 October 2015
Accepted 9 October 2015
Available online 23 October 2015

Keywords:
Source term estimation
Bayesian inference
Monte-Carlo techniques
Adaptive multiple importance sampling

ABSTRACT

In the eventuality of an accidental or intentional atmospheric release, the reconstruction of the source term using measurements from a set of sensors is an important and challenging inverse problem. A rapid and accurate estimation of the source allows faster and more efficient action for first-response teams, in addition to providing better damage assessment.

This paper presents a Bayesian probabilistic approach to estimate the location and the temporal emission profile of a pointwise source. The release rate is evaluated analytically by using a Gaussian assumption on its prior distribution, and is enhanced with a positivity constraint to improve the estimation. The source location is obtained by the means of an advanced iterative Monte-Carlo technique called Adaptive Multiple Importance Sampling (AMIS), which uses a recycling process at each iteration to accelerate its convergence.

The proposed methodology is tested using synthetic and real concentration data in the framework of the Fusion Field Trials 2007 (FFT-07) experiment. The quality of the obtained results is comparable to those coming from the Markov Chain Monte Carlo (MCMC) algorithm, a popular Bayesian method used for source estimation. Moreover, the adaptive processing of the AMIS provides a better sampling efficiency by reusing all the generated samples.

1. Introduction

The threat of chemical, radiological, biological, and nuclear (CRBN) releases in the atmosphere is a key issue. Such incidents may be the consequence of intentional releases using non-conventional methods in order to create panic. The origin of these releases can also be accidental, for example given a leak of hazardous material on an industrial site. Either way, the development of tools to reconstruct the source is of utmost importance for the population safety as well as for the efficiency of the first-response teams action.

Scientifically speaking, the question of source term estimation (STE) is a challenging inverse problem, due to its ill-posed nature. First, there is the issue of the non-uniqueness of the source reconstruction solution: for a given set of concentration measurements obtained by the sensors, there may be several possible

* Corresponding author. CEA-DAM, DIF, F-91297 Arpajon, France.
E-mail address: harizo.rajaona@cea.fr (H. Rajaona).

http://dx.doi.org/10.1016/j.atmosenv.2015.10.026
1352-2310/

source terms that would fit the observations. This problem is aggravated by the fact that the observations could occasionally be non-representative of the reality, because of the errors introduced in the acquisition process such as the sensor noise, or the potential averaging of the measurements done by the receptors. Next comes the fact that the process of estimating the source may easily be jeopardized, due to the physical mechanisms it relies on: the micro-meteorological parameters involve complex processes which may provide quite different outputs in certain conditions, should those parameters vary in small or large amplitudes.

To solve the problem of STE, several methods have been developed, using different approaches. The most physically intuitive one is the use of *inverse transport*, which consists in running an atmospheric dispersion model backward in time, from the observation times to the release times (Robertson and Langner, 1998; Pudykiewicz, 1998). The backward model relies on an adjoint transport equation, which is derived by using the principle of time-symmetry employed in atmospheric transport (Hourdin and Talagrand, 2006): switching from the forward model requires changing the sign of each term but the diffusion in the transport equations. Here, the ill-posed aspect of the STE problem is treated by resorting to a regularization process, allowing the construction of a well-posed unique solution by minimizing a cost function. The idea behind this is to penalize the non-desired solutions of the STE problem by applying constraints on them, and to perform a goodness-of-fit measure between the predicted concentrations given by the model and the real observations. There are several ways to optimize the cost function: one approach is to use genetic algorithms, based on evolutionary computation, that runs a selection-and-mutation process to a population of candidates in order to provide the optimal solution (see Haupt et al., 2007; Allen et al., 2007; Cervone et al., 2010; Cervone and Franzese, 2011; Annunzio et al., 2012 for examples of genetic algorithms used in the STE context). Several alternatives are available and have been studied, such as the concept of illumination (Issartel, 2005), or principles relying on information theory, such as the maximum entropy on the mean (Bocquet, 2005).

Inverse modelling can also be viewed as a Bayesian problem (Winiarek et al., 2011) where the expression of the cost function is established using Bayesian inference, introducing probabilistic considerations in the problem in order to obtain a point estimate of the source parameters. One Bayesian inference strength is indeed to be able to estimate the uncertainty factor regarding the estimation. Another way of exploiting the Bayesian approach consists in seeking not just for one optimal solution, but obtaining the probability distribution of the source characteristics using the Bayesian paradigm: in this case, the source is defined by a set of parameters, which are the quantities of interest. By the means of stochastic sampling, the posterior probability distribution of these parameters is evaluated in order to fully describe the parameters of the source and the uncertainty on them. The goal of STE is then to look for the most probable parameters for the source in terms of posterior probability. The most frequently used tool for sampling from this posterior distribution in STE problems is the family of Markov Chain Monte Carlo (MCMC) algorithms. In Keats et al., 2007, MCMC is used to determine the source parameters in a complex urban environment. In Senocak et al., 2008, the MCMC procedure is embedded into a reconstruction tool that also assesses the wind field parameters of the dispersion model. In Chow et al., 2008, MCMC is implemented alongside a *Computational Fluid Dynamics* (CFD) model that aims at improving the accuracy of the physical model. In Delle Monache et al., 2008, a parallel processing scheme using MCMC is built in order to reconstruct the source term of the Algeciras incident. In Yee et al., 2014, MCMC is used to reconstruct the location and the emission rate of a medical isotope production facility using activity concentration measurements obtained from the International Monitoring System radiological network. Overall, MCMC methods have proven to give good results, and their use has extended to other similar purposes, such as water pollution detection (Hazart et al., 2014) or natural gas seeps mapping (Hirst et al., 2013).

The method we introduce in this paper offers an alternative way of tackle this Bayesian inference problem, by developing an adaptive scheme based on the principle of Importance Sampling (IS). This method allows to enhance the estimation process by learning from the previous sampling rounds through a recycling process, thus accelerating the convergence and reconstructing faster the source term parameters. In Section 2, we describe the Bayesian framework and the formalism of our study. In Section 3, we develop the foundations of the methodology around the AMIS algorithm for source location purposes. Finally, in Section 4, we implement our solution to the STE problem using a framework derived from the Fusion Field Trials 2007 experiment and compare its behaviour to the MCMC, before concluding.

2. Bayesian modelling

In this paper, we are interested in estimating the unknown characteristics $\boldsymbol{\theta}$ of a source, given the measurements, $\boldsymbol{d}$, obtained from all the sensors deployed in the field. More specifically, a Bayesian solution will be designed for the inference in order to take into account all the statistical information given by our problem. As a consequence, the quantity of interest is the posterior distribution given by the Bayes's rule as:

$$p(\boldsymbol{\theta}|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{d})} \tag{1}$$

As we can see the main ingredients of Bayesian analysis are both the prior distribution, $p(\boldsymbol{\theta})$, and the likelihood distribution, $p(\boldsymbol{d}|\boldsymbol{\theta})$. The prior distribution represents the beliefs about the unknown state before obtaining any observations. On the other hand, the likelihood distribution gives the probability of obtaining the data given a certain set of parameters values. $p(\boldsymbol{d})$ represents the marginal likelihood and acts as the normalizing constant of the product of the prior and likelihood distributions in order to obtain the posterior distribution. Let us now specify both the likelihood and the prior distributions for the problem considered in this paper.

2.1. The likelihood model

In this work, we consider a point-wise and static source represented by $\boldsymbol{\theta}=(\boldsymbol{x}_s,\boldsymbol{q})$ where $\boldsymbol{x}_s=(x_s,y_s)$ is the spatial position of the source and $\boldsymbol{q}$ is the release rate vector resulting from the discretization of the plausible emission time interval into T_s time steps, $\{\Delta t_n\}_{n=1}^{T_s}$, in order to take into account the possible time dependence aspect of the source.

Let us assume the concentration is measured at a finite number of time points by a set of N_c sensors deployed in the field. The observed concentration $d_{i,j}$ acquired by the i-th sensor c_i at location $\boldsymbol{x}_{c_i}$ and time t_j, where $i=1,\ldots,N_c$ and $j=1,\ldots,T_c$ with T_c the number of time samples for each sensor, is modelled by:

$$d_{i,j} = \sum_{n=1}^{T_s} q_n \mathbf{C}_{i,j}(\boldsymbol{x}_s, \Delta t_n) + \varepsilon_{i,j} \tag{2}$$

The first term in the right-hand side of this expression corresponds to the mean concentration resulting from the superposition of the T_s releases on the different time steps $\{\Delta t_n\}_{n=1}^{T_s}$ weighted by the actual emission rates $\{q_n\}_{n=1}^{T_s}$ of the source. $\mathbf{C}_{i,j}(\boldsymbol{x}_s,\Delta t_n)$